Journal of Econometrics 186 (2015) 19–31

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Inference on higher-order spatial autoregressive models with increasingly many parameters

Abhimanyu Gupta [a], Peter M. Robinson [b,*]

[a] Department of Economics, University of Essex, UK

[b] Department of Economics, London School of Economics, UK

ARTICLE INFO

Article history:
Received 2 December 2013
Received in revised form
23 August 2014
Accepted 19 December 2014
Available online 7 January 2015

JEL classification:
C21
C31
C33
C36

Keywords:
Spatial autoregression
Increasingly many parameters
Central limit theorem
Rate of convergence
Spatial panel data

ABSTRACT

This paper develops consistency and asymptotic normality of parameter estimates for a higher-order spatial autoregressive model whose order, and number of regressors, are allowed to approach infinity slowly with sample size. Both least squares and instrumental variables estimates are examined, and the permissible rate of growth of the dimension of the parameter space relative to sample size is studied. Besides allowing the number of parameters to increase with the data, this has the advantage of accommodating some asymptotic regimes that are suggested by certain spatial settings, several of which are discussed. A small empirical example is also included, and a Monte Carlo study analyses various implications of the theory in finite samples.

1. Introduction

Correlation in cross-sectional data poses considerable challenges, complicating both modelling and statistical inference. When information on geographical locations is available, it may be possible to extend models developed for time series data. However, when locations are irregularly-spaced serious difficulties arise, and frequently only information on economic (not necessarily geographic) distances is available. Spatial autoregressive (SAR) models, due to Cliff and Ord (1973), have become widely used in this setting. Given a sample of size n, these employ a known $n \times n$ spatial weights matrix whose (i, j)th element is inversely related to some measure of economic distance between units i and j. The elements may also be binary, for instance taking equal values 1 when two units are contiguous and 0 otherwise, but many other specifications are possible.

To be specific, for an $n \times 1$ vector of observations y_n, an $n \times k$ matrix of regressors X_n and $n \times n$ weight matrices $W_{in}, i = 1, \ldots, p$, it is assumed that there exist unknown scalars $\lambda_1, \lambda_2, \ldots, \lambda_p$ and an unknown $k \times 1$ vector β such that

$$y_n = \sum_{i=1}^{p} \lambda_i W_{in} y_n + X_n \beta + U_n, \tag{1.1}$$

where U_n is an $n \times 1$ vector of disturbances. This model, allowing $p > 1$, has been studied by, e.g., Blommestein (1983), Huang (1984), Huang and Ahn (1992), Anselin (2001), Lee and Liu (2010) and Badinger and Egger (2013). In this paper we will refer to the above as the SAR model while the SAR model without X_n will be the *pure* SAR model.

Weight matrices need not be symmetric and can contain negative elements, but their diagonal elements are zero, and they are frequently row-normalized such that each row sums to 1. If W_{in} has non-negative elements, this implies that its (j, l)th element can be interpreted as $w_{jl,in} = d_{jl,in} / \sum_{h=1}^{n} d_{jh,in}$, where $d_{jl,in}$ measures inverse distance between units j and l. Thus elements of the W_{in} are allowed to depend on n, so those of y_n form a triangular array. Since X_n may also depend on spatial weights, we also allow its elements to depend on n. See e.g. Arbia (2006) for a review of spatial autoregressions.

* Corresponding author. Tel.: +44 20 7955 7516; fax: +44 20 7955 6592.
E-mail addresses: a.gupta@essex.ac.uk (A. Gupta), p.m.robinson@lse.ac.uk (P.M. Robinson).

http://dx.doi.org/10.1016/j.jeconom.2014.12.008

By far the most popular version of (1.1) takes $p = 1$, when we write

$$y_n = \lambda W_n y_n + X_n \beta + U_n. \quad (1.2)$$

Due to the spatially lagged y_n on the right, ordinary least squares (OLS) estimation of λ and β is problematic, but Lee (2002) showed that under suitable conditions OLS can be consistent, and asymptotically normal and efficient. In particular, for a divergent positive sequence h_n that is bounded away from zero uniformly in n, consistency follows if $w_{ij,n} = \mathscr{O}\left(h_n^{-1}\right)$ and asymptotic normality if also $n^{\frac{1}{2}}/h_n \to 0$ as $n \to \infty$.

Instrumental variables (IV) estimation (see Kelejian and Prucha (1998)) is $n^{\frac{1}{2}}$-consistent under less restrictive conditions than OLS, but inefficient. On the other hand, it is computationally simpler than procedures which may have better statistical properties, such as generalized method of moments (Kelejian and Prucha, 1999; Lee and Liu, 2010), optimal IV (Lee, 2003), Gaussian pseudo maximum likelihood (Lee, 2004), and adaptive estimation (Robinson, 2010). Additionally, desirable asymptotic properties of OLS and IV require X_n to contain at least one non-intercept regressor.

In this paper we allow the spatial lag order p in (1.1) and the number of regressors k to increase slowly with n, as opposed to being fixed. This scheme reflects the practical reality that the richness of a parametric model often deepens with sample size, and has been explored previously in various settings. The model studied in this paper is defined explicitly in the next section.

Higher-order SAR models (such that $p > 1$ in (1.1)) raise serious identifiability problems, which become more acute if p is allowed to increase with n. Our assumptions for consistency and asymptotic normality imply identifiability, but the practitioner needs to be aware of what choices of the W_{in} can potentially afford it. It is obvious that no W_{in} can be a linear combination of the others, but this property is far from sufficient (see e.g. Anselin (2001)). One particular class of W_{in} will transparently avoid identifiability problems, in particular ones with 'single nonzero diagonal block structure'. To define these, denote by V_n an $n \times n$ block diagonal matrix with ith block V_{in}, a $m_i \times m_i$ matrix, where $\sum_{i=1}^{p} m_i = n$, and for $i = 1, \ldots, p$ obtain W_{in} from V_n by replacing each $V_{jn}, j \neq i$, by a matrix of zeros. Thus $V_n = \sum_{i=1}^{p} W_{in}$.

This structure can be thought of as extending a choice of W_n in (1.2) suggested by Case (1991, 1992), where each of p districts contains m farmers, so $n = mp$, and there is interdistrict independence, implying a block diagonal W_n, and also homogeneous within-district reactions, so

$$W_n = I_p \otimes B_m, \quad \text{with } B_m = (m-1)^{-1}\left(l_m l_m' - I_m\right), \quad (1.3)$$

where prime denotes transposition, l_m is the m-dimensional vector of ones $(1, \ldots, 1)'$, I_m is the m-dimensional identity matrix and $\otimes$ denotes Kronecker product. For (1.1) we might then take the W_{in} to have single nonzero diagonal block structure such that

$$m_i = m, \qquad V_{in} = B_m, \quad i = 1, \ldots, p, \quad (1.4)$$

allowing SAR coefficients to vary across districts. In asymptotic theory for (1.1) using (1.3), p is sometimes allowed to diverge with n, and if that happens with (1.4) the number of possibly distinct λ_i likewise increases. Of course (1.4) might then be extended to allow unequal m_i and V_{in} across i, and V_{in} with more complex structure. As in the statistical literature on regression models, we also allow k to increase slowly with n. This theme has been pursued in a variety of models (see e.g. Huber, 1973; Berk, 1974; Robinson, 1979, 2003; Portnoy, 1984, 1985), but not previously with SAR models. Pinkse et al. (2002) consider nonparametric series estimation of a model with spatial weights determined by an unknown function of economic distances.

The following section introduces some assumptions that are basic to our theoretical results. Further assumptions, and theorems, for the consistency and asymptotic normality of IV and OLS estimates are presented in Sections 3 and 4 respectively. In Section 5 we consider some illustrations, followed by an empirical example in Section 6. We conduct a Monte Carlo study in Section 7, while Section 8 concludes. Proofs are in Appendices A and B

2. Model and basic assumptions

We rewrite (1.1) to stress the possible dependence of the parameter dimension, and the parameters themselves, on n and also introduce an endogenous regression component:

$$y_n = \sum_{i=1}^{p_n} \lambda_{in} W_{in} y_n + H_n \gamma_{(n)} + X_n \beta_{(n)} + U_n, \quad (2.1)$$

where $\lambda_{(n)} = \left(\lambda_{1n}, \ldots, \lambda_{p_n n}\right)'$, $\gamma_{(n)} = \left(\gamma_{1n}, \ldots, \gamma_{s_n n}\right)'$ and $\beta_{(n)} = \left(\beta_{1n}, \ldots, \beta_{k_n n}\right)'$. We may write (2.1) as

$$S_n y_n = H_n \gamma_{(n)} + X_n \beta_{(n)} + U_n, \quad (2.2)$$

where $S_n = I_n - \sum_{i=1}^{p_n} \lambda_{in} W_{in}$, or equivalently $y_n = R_n \lambda_{(n)} + H_n \gamma_{(n)} + X_n \beta_{(n)} + U_n$ with $R_n = (W_{1n} y_n, \ldots, W_{p_n n} y_n)$. X_n will be taken to comprise exogenous elements, while s_n can also diverge and the elements of H_n are allowed to be correlated with u_n. We now introduce some basic assumptions.

Assumption 1. $U_n = (u_1, \ldots, u_n)'$ has iid elements with zero mean and finite variance σ^2.

Assumption 2. For $i = 1, \ldots, p_n$, the diagonal elements of each W_{in} are zero and the off-diagonal elements of W_{in} are uniformly $\mathscr{O}\left(h_n^{-1}\right)$, where h_n is some positive sequence which is bounded away from zero and which may be bounded or divergent, with $n/h_n \to \infty$ as $n \to \infty$.

Different h_{in} sequences for each of the W_{in} may be used. However for OLS estimation, even for fixed p, Lee (2002) demonstrated that consistency requires divergence so that $\min_{i=1,\ldots,p_n} h_{in} \to \infty$ must be assumed. He also provided a detailed discussion of this assumption. In IV estimation, any mixture of bounded and divergent h_{in} sequences may be employed. Boundedness away from zero is crucial as even consistency of the error variance estimate based on IV residuals may fail if this does not hold, Indeed, in the 'farmer-district' setting discussed in the previous section, $h_n = m - 1$.

Assumption 3. S_n is non-singular for sufficiently large n.

This assumption ensures that (2.2) has a solution for y_n. In certain special cases such as the farmer-district setting presented above, a sufficient condition can be provided for Assumption 3. For any $s \times q$ matrix $A = \left[a_{ij}\right]$ we define $\|A\|_R = \max_{i=1,\ldots,s} \sum_{j=1}^{q} \left|a_{ij}\right|$, the maximum absolute row-sum norm. The proof of the following can be found in the Appendix.

Proposition 2.1. *When for each $i = 1, \ldots, p_n$, $\|W_{in}\|_R \leq 1$ and each W_{in} has a single nonzero diagonal block structure, a sufficient condition for invertibility of S_n is that $|\lambda_{in}| < 1$, $i = 1, \ldots, p_n$.*

Assumption 4. $\left\|S_n^{-1}\right\|_R$, $\left\|S_n'^{-1}\right\|_R$, $\|W_{in}\|_R$ and $\left\|W_{in}'\right\|_R$ are bounded uniformly in n and i, $i = 1, \ldots, p_n$, for sufficiently large n.

This assumption is standard, the parts pertaining to S_n^{-1} ensuring that the spatial correlation is curtailed to a manageable degree because the covariance matrix of y_n conditional on the regressors is $\sigma^2 S_n^{-1} S_n'^{-1}$, while those for the W_{in} are satisfied trivially if one unit is assumed to be a 'neighbour' of only a finite number of other units, and more generally satisfied if, for each i, the elements of W_{in} decline fast enough with n, as is natural if they are inverse distances.